Measuring Similarity between Brands using Followers' Post in Social Media

Yiwei Zhang
zhangyiwei@hal.t.u-tokyo.ac.jp
The University of Tokyo

Xueting Wang
xt_wang@hal.t.u-tokyo.ac.jp
The University of Tokyo

Yoshiaki Sakai
sakai.yoshiaki@geomarketing.co.jp
Geomarketing Co.,Ltd.

Toshihiko Yamasaki
yamasaki@hal.t.u-tokyo.ac.jp
The University of Tokyo

ABSTRACT

In this paper, we propose a new measure to estimate the similarity between brands via posts of brands' followers on social network services (SNS). Our method was developed with the intention of exploring the brands that customers are likely to jointly purchase. Nowadays, brands use social media for targeted advertising because influencing users' preferences can greatly affect the trends in sales. We assume that data on SNS allows us to make quantitative comparisons between brands. Our proposed algorithm analyzes the daily photos and hashtags posted by each brand's followers. By clustering them and converting them to histograms, we can calculate the similarity between brands. We evaluated our proposed algorithm with purchase logs, credit card information, and answers to the questionnaires. The experimental results show that the purchase data maintained by a mall or a credit card company can predict the co-purchase very well, but not the customer's willingness to buy products of new brands. On the other hand, our method can predict the users' interest on brands with a correlation value over 0.53, which is pretty high considering that such interest to brands are high subjective and individual dependent.

CCS CONCEPTS

• **Information systems** → **Social advertising**; *Social networking sites*; *Content ranking*; *Social recommendation.*

KEYWORDS

Social Media, Hashtag, Brands, Computational Marketing

1 INTRODUCTION

Billions of photographs are uploaded to the Internet every day through various photo & video sharing services. Social media is already a part of our everyday life. The rapidly increasing popularity of social media has given rise to new marketing approaches. Social media marketing is the use of social media platforms and websites to promote a product or service [11]. Companies make use of platforms such as Twitter[1] and Instagram[2] to reach a much wider spectrum of audience than that of traditional advertisements such as newspaper, television, and radio. By 2018, 91% of the Fortune 500 companies were actively using Twitter, and 63% of them had corporate Instagram accounts[3]. With these newfound trends, we can use data from social media to analyze customers' preferences and the implicit relationship between brands. Understanding the relationship between brands, products, and people can have a profound effect on marketing. For example, when a brand wants to find celebrities for endorsements, it is important to consider whether his/her followers are potential customers for its products. Similarly, when a shopping mall is in the process of choosing the brands it wishes to carry, considering the specific combinations of brands that have the potential to attract more customers is a lucrative proposition. However, such design or decision making has been done by only experts with sense of trend so far.

Co-purchasing history from credit cards or point cards in a shopping mall can be analyzed to measure the similarity between brands. Unfortunately, there are some demerits of this method. Some customers, especially teenagers, may not own a credit card. Secondly, people prefer using cash for small purchase amounts. Consequently, brands that offer products in the low price range may seldom appear in the purchasing history. Many shopping malls have their own point card systems, which allow customers to earn points every time they make a purchase from one of the franchises. But these analyses are limited by the number of brands existing at the location of the purchase. Namely, similarity between brands that are not co-located in the same region can not be calculated.

Because of these reasons, we chose social network data to measure the similarity between brands. Unlike the prerequisites for credit cards or point cards, social media has no age or income requirement, also there are no limitations to the number of brands being advertised by marketers or users. One might think that SNS penetration is lower among elder generation and therefore, it is

---

[1]https://twitter.com
[2]https://www.instagram.com
[3]https://www.umassd.edu/cmr/social-media-research/2018-fortune-500

hard to get information from them. But according to the investigation[4], 64% of Americans in the age group 50-64 use SNS actively as well as 37% in the age group of 65 years and above.

We briefly summarize the main contributions and findings of this study:

- We have presented an approach to predict the similarity between brands using tags and images posted by the brands' followers. We show that such off-the-shelf multimedia technologies, though not technically novel, can be a new marketing tool.
- We have created an Instagram brand datasets including 233,000 unique users and their most recent 100 posts for brand similarity measurement. We release Instagram dataset to the public[5]. We have also created two evaluation datasets based on the purchasing histories of two real-world customers for results comparison.
- We have conducted user studies by questionnaires to obtain the users' co-purchasing, interest, and knowledge tendencies. It provides an additional angle of evaluation besides point card and credit card purchasing histories.

2 RELATED WORKS

2.1 Similarity Measurement

There are many related studies on measuring latent similarity with multimedia data. Item2Vec [2] measured similarity between item for recommendation. They considered items purchased by one user as a positive example then used a skip-gram model to produce embeddings. Eisner et al. used a similar method to convert an emoji to a vector then measured similarity between emoji [10]. [16] measured latent similarity between social images by metric learning. Wang et al. used graph embedding techniques to find similar items [21]. Hsiao et al. used the topic models to learn style-coherent representations for fashion images retrieval [15].

2.2 Brand Analysis using Social Media

Researches on brands are increasingly focused on the content on social media. Goh et al. showed that engagement in social media brand communities leads to a positive increase in sales [13]. In a study of understanding brands' perceptions, Culotta et al. used the data on Twitter to predict how consumers perceive brands' attributes including eco-friendliness, nutrition, and luxury [7]. Several researchers analyzed the popularity of brands and their social media posts. De Vries et al. reported that brand posts might be popular due to several cues such as vividness, interactivity, and valence of comments [8]. Mazloom et al. investigated similar indicators with text, image and social features on Instagram [17].

2.3 Content-based Recommendation

Recommendation is another topic of interest in the study of social media, both commercially and academically. Besides widely used collaborative filtering methods [6], another common approach is content-based recommendation [18]. Apart from users-item interactions, content-based methods analyze information such as user
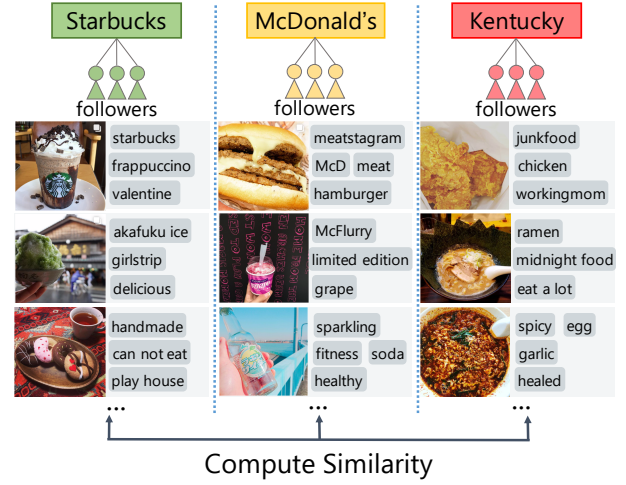
---

[4]https://www.pewinternet.org/fact-sheet/social-media
[5]https://github.com/yiwei51/brand_dataset

Figure 1: Example posts from brand followers.

profiles and item relations [22]. Gelli et al. presented a study on image recommendations for brands. They recommended images which were consistent with a brand's style, by learning features from images uploaded by these brands [12].

2.4 Brand Similarity

The pairwise similarity between brands was measured previously by Bijmolt et al. [3]. They used questionnaires to ask people regarding the likeness of two products to identify the similarities between the products of various brands. The limitation of this method was that it required time and monetary investment to collect sufficient data. In [23], they aimed to find similar brands in the same category using customers' reviews, while their approach is not suitable for finding the relationship between brands in diverse product categories.

In [20], the researchers predicted user's personality based on his or her profile picture on Facebook. We also believe that images and hashtags posted by users are indicative of their personalities and users' choice of brands represents their preferences. Different from [3] and [23], our proposed method could find the relationship between two diverse product categories without tremendous time and monetary investment.

3 DATASET CONSTRUCTION

Instagram is a social photo sharing service. We chose to collect posts from Instagram because many brands have created official accounts for promoting the brand image. We show posts from the followers of three brands separately in Figure 1. Though these posts are not from the brands' official accounts, we can see that followers of different brands have different styles. Their posts indicate the everyday activities of these followers, which, we believe, can be utilized to represent the similarity between the brands. We chose to use followers' posts instead of brands' posts because brands in the same category may have greater similarity based on their posts. For example, a fashion brand will be more similar to another fashion brand than to a cosmetic brand. However, finding the relationship between two diverse product categories is of greater interest to us.

Table 1: Summary of datasets used in this paper.

Dataset	# Brands	# Instagram users (1,000 followers/brand)	Data source	# Transactions	# Customers	Period of record
Pointcard100	100	100,000	point card used in a mall	1,759,098	241,937	2015.04.01 - 2016.03.31
Creditcard81	81	81,000	a national credit card brand	9,393,546	1,236,521	2016.01.01 - 2017.12.31
Popular108	108	108,000	Instagram	-	-	-
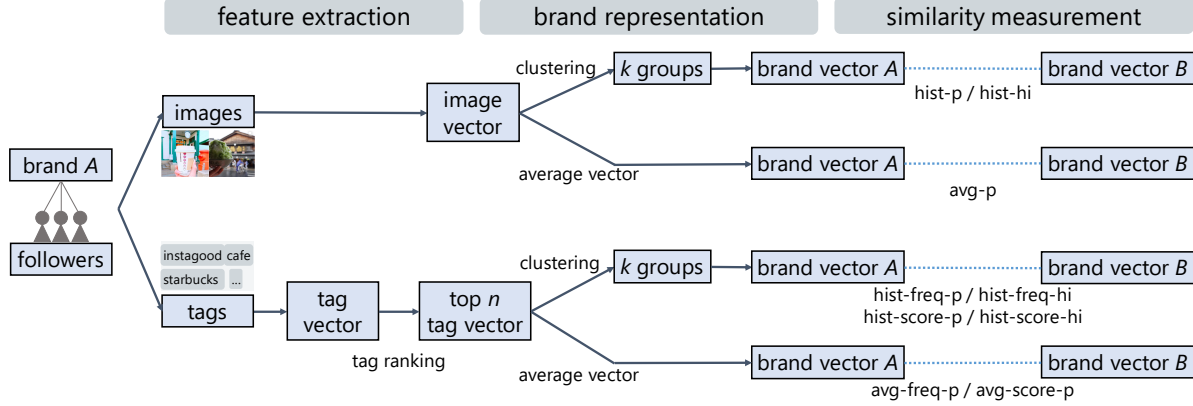
Figure 2: The flowchart for the feature extraction, brand representation and similarity measurement processes.

The problem of measuring the pairwise similarity between brands transforms to the problem of measuring the pairwise similarity between their followers. We randomly selected 1,000 followers per brand and used their posts to represent this brand. We chose to associate a unique user to just one brand. We collected data on these followers by web crawling their Instagram posts.

We created three datasets in our study. The difference between the datasets is shown in Table 1. Pointcard100 and Creditcard81 datasets are created for the evaluation. The Pointcard100 dataset included 100 brands that was obtained from customers' purchasing histories at a shopping mall in Japan. The Creditcard81 dataset included 81 brands and comprised customers' purchasing histories maintained by a credit card company in Japan. They were both anonymized before processing. Both datasets included information on products of specific brands purchased by the users. The Popular108 dataset included the official accounts of 108 of the most popular companies in Japan. The number of followers in Instagram was directly used to express the popularity.

4 PROPOSED METHODS

We present a new scale for measuring the similarity between brands by analyzing posts from brand followers. A flowchart for the proposed methods consisting of the feature extraction, brand representation and similarity measurement processes is depicted in Figure 2.

4.1 Feature Extraction

4.1.1 Image Feature. For each brand, we used the 10 most recent photographs posted by 1,000 users to represent the brand. We used a 50-layer ResNet [14] pretrained on ImageNet [9] to extract the image features. In this step, we transform each image into a feature vector of column size 2048.

4.1.2 Tag Feature. To extract efficient tag features, we took two steps: (1) tag embedding to covert each tag into a vector representation; (2) tag ranking to select the top ranked tags as the typical tag feature for brand representation.

Tag embedding process is inspired by the work Item2Vec [2]. It produces embedding for items in a latent space using Word2Vec. We considered tags belonging to one image as words in one sentence. Then, we used *fasttext*™ to convert each tag to a 100-dimension vector [5].

In the tag ranking, we select the top 3,000 tags to represent the brand. The first method is to rank by frequency and the second method is to rank by tag score.

Rank by Frequency: In this method, we rank tags by the number of users who have used it. In [4], the researchers observed that a large number of users use the same tags to every photo. Therefore, we decided to rank the tags by the number of users who have used the tag at least once. In other words, regardless of the number of times the user has used the tag, we only counted the tag once for one user.

Rank by Tag Score: Instead of utilizing the number of users who have used the tag at least once, we apply term frequency–inverse document frequency (TF–IDF) [19] for selecting the tags. TF–IDF is a numerical statistic, which indicates the importance of a word to a document in a collection or corpus [1]. In our tag ranking algorithm, we considered a brand as a document, and tags belong to a brand as words inside a document. We utilized the number of users who have used the tag at least once to represent the TF. The TF score is higher if more people have used the tag. In contrast, we consider that the IDF score would be lower if the tag has been used for other brands frequently. The score for the tag is calculated

using the following equation:

$$tag\_score(t_i, b_j) = tf(i,j) \times idf(i), \qquad (1)$$

$$tf(i,j) = \sum_k \frac{n_{i,j}}{n_{k,j}}, \qquad (2)$$

$$idf(i) = \log \frac{B}{\sum_j c_{i,j}} + 1, \qquad (3)$$

$$c_{i,j} = \begin{cases} 1 & r_{i,j} \le l \\ 0 & r_{i,j} > l, \end{cases} \qquad (4)$$

where $n_{i,j}$ is the number of users that have used the tag t_i in brand b_j, and B is the number of brands. $r_{i,j}$ is the frequency ranking of tag t_i in the brand b_j, $c_{i,j}$ is 1 if the tag t_i appears in the top l tags of brand b_j, else $c_{i,j}$ is 0. We set $l = 1,000$ in this experiment.

4.2 Brand Representation

After converting the images and tags into vectors, we addressed the problem of representing a brand using these vectors. In the following section, we introduce several ways to handle image and tag features, including histogram-based methods and average vector based methods, and compare their performances.

Histogram-based Method: We used mini-batch K-means to cluster features after feature extraction to form a bag of features. We have tried using K from 10 to 10,000 and employed the best one. After clustering, we use the number of images/tags in each cluster as the brand vector.

Average Vector based Method: In the case of images, we average the image vectors directly and use it as the brand vector. In the case of tags, after tag ranking, we select top 3,000 tags and average their tag vectors, then we use this averaged vector as the brand vector.

4.3 Similarity Measurement

We considered two ways for histogram-based methods in similarity measurement. The first is the Pearson correlation and the second is histogram intersection similarity. For the average vector based methods, we only considered the Pearson correlation. We calculated similarity between each pair of brand vectors. Figure 3 shows the visualization of the brand relationships based on the histogram method of ranking by the tag score. We can easily find relationship between brands from this figure. For example, for those who like "Starbucks," they might also be interested in fashion brands "titivate."

5 EVALUATION

5.1 Stability

Because our proposed methods are based on 1,000 followers randomly selected for each brand, we conducted an experiment to evaluate the stability of our proposed methods. For each brand, we randomly split the 1,000 followers into two groups for 5 times and then computed the similarity between the brands for each group. The average Spearman's ranking correlation coefficient based on the two groups was calculated as 0.98. It indicates that our proposed methods are stable.

We also conducted an experiment to show that less number of followers might be enough for our purpose. We randomly selected a small group of followers from each brand for 5 times and then computed the similarity between using a small group of followers

and using 1,000 followers. We found that the average Spearman's ranking correlation coefficient was calculated as 0.93 when using 100 followers per brand, and 0.88 when using 50 followers per brand. We think 100 followers might be enough for our purpose.

5.2 Evaluation using Co-Purchasing Data

5.2.1 Method. To prove that our results have implications for real-world business, we compared our results with the purchasing histories of customers in the Pointcard100 and Creditcard81 datasets. We were able to create a brand-user matrix M based on the co-purchasing data. $M_{b_i u_k} = t$ indicates that user u_k purchased brand b_i t times. Then, we calculated the Pearson correlation coefficient between each pair of brands as similarity score. For example, the similarity score between brand b_i and b_j was calculated using vector M_{b_i} and M_{b_j}. Then, we compared the results based on our proposed methods with the results calculated from the co-purchasing data using the Spearman's ranking correlation.

5.2.2 Results of Evaluation using Co-Purchasing Data. Table 2 shows the Spearman's ranking correlation coefficient for brand similarity based on our proposed methods and based on co-purchasing results from the point card and credit card usage. In all tables, *hist* stands for the histogram-based methods; *avg* stands for the average vector methods; *p* stands for using the Pearson correlation; *hi* stands for using histogram intersection similarity; *freq* stands for ranking by frequency; *score* stands for ranking by tag score. For histogram-based methods, results changed a little when choosing different K and we empirically found that using $K = 500$ has better result in our proposed methods. We also compared our method with *brand_img* by [12]. Their method used images posted by the brand as the input and produced a brand vector that aimed to discovered images that match the brand style. We used their implementation to train the brand vector then calculate the Pearson correlation between each pair of brands as similarity score.

In the case of images, the average vector based method yields better results than the histogram-based methods. Also, proposed methods show better performance than *brand_img* [12], proving that using followers' posts is better. We could see that the tag feature works better than the image feature when compared with the co-purchasing results. When using the histogram-based methods, histogram intersection similarity yields better results than the Pearson correlation. With regard to the average vector based methods, ranking by tag score gives better results than ranking by frequency. The highest Spearman's ranking correlation coefficient achieves 0.50, which is pretty high considering that this result was calculated among different groups of users (point/credit card users and brand followers are totally different persons). The correlation coefficients of the Creditcard81 is lower than those of Pointcard100 because the users of the credit card are distributed all over the country and it is not guaranteed that all the brands have their shops within their shopping territory.

5.3 Evaluation using Questionnaires

5.3.1 Method. To further evaluate our results, we used questionnaires to ask users regarding their awareness of certain brands and their past purchases of the products of those brands through a crowdsourcing service. Two separate questionnaires were asked
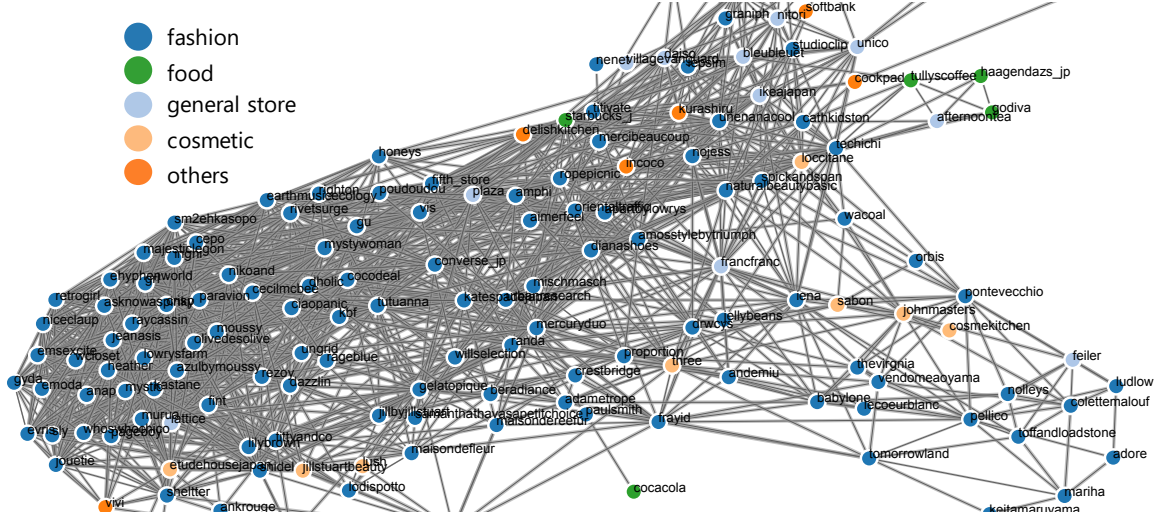
Figure 3: Visualization of the brand relationships (full version of the visualization[6]).

Table 2: Spearman's ranking correlation coefficient between proposed methods and co-purchasing results.

		Co-purchasing result	
		Pointcard100	Creditcard81
Image	brand_img [12]	0.22	0.19
	hist-p	0.33	0.34
	hist-hi	0.32	0.33
	avg-p	**0.37**	**0.37**
Tag	hist-freq-p	0.44	0.4
	hist-freq-hi	**0.50**	0.43
	hist-score-p	0.42	0.41
	hist-score-hi	0.48	0.43
	avg-freq-p	0.44	0.43
	avg-score-p	0.45	**0.44**

for the Pointcard100 and Creditcard81 datasets. Nine hundred people answered the questionnaire for Pointcard100 dataset, and 890 people answered the questionnaire for Creditcard81 dataset. For both questionnaires, we selected the top 50 brands from the dataset according to their number of followers on Instagram. The two questionnaires were similar. We asked people three questions related to each brand.

- *Have you purchased this brand by yourself before?*
- *Are you interested in this brand?*
- *Do you know this brand?*

We obtained the users' co-purchasing, interest and knowledge tendencies from the responses to the questionnaires. Similar to the purchasing data in the previous section, we created three brand-user matrices, which represented the co-purchasing, interest and knowledge tendencies separately. For example, for the first question, the response $M_{b_i u_k} = 1$ indicated that user u_k had bought brand b_i before, whereas the response $M_{b_i u_k} = 0$ indicated that user u_k had never bought brand b_i before. Then, we calculated the Pearson correlation coefficient between each pair of brands.

[6]https://yiwei51.github.io/brand_visualization

5.3.2 Results of Evaluation using Questionnaires. Table 3 shows the results for the two datasets. For Pointcard100 dataset, we can see that the Spearman's ranking correlation coefficient between the co-purchasing results obtained from the point card data and from the questionnaires is 0.52, which is high considering that co-purchasing activity is strongly affected by locality, personality, etc. The co-purchasing results obtained from the point card data have higher correlation than the results calculated from our proposed methods. However, comparison of the customer interest obtained from the questionnaires and that from the knowledge results shows that the proposed method based on histogram has the highest Spearman's ranking correlation coefficient. The highest correlation with the customer interest results obtained from the questionnaires is 0.53, whereas the highest correlation with the knowledge results obtained from the questionnaires is 0.50. For Creditcard81 dataset, when compared with the customer co-purchasing, interest and knowledge results obtained from the questionnaires, the proposed methods based on the tags show stronger correlation than the co-purchasing results obtained from the credit card data.

From these results, it can be said that we have grasped the tendency of customer preferences. Our proposed methods can predict customers' interests more accurately than the predictions made from the customer's purchasing history. We considered the reason for this result might be related to age or finances. For example, women in their early twenties might be unable to purchase some high-end fashion brands because of age-related and insufficient finances, but they might be interested in these brands to prepare for their late twenties or thirties. We believe that our proposed methods could reflect a customer's future purchasing plan.

6 APPLICATIONS

We have introduced that different brands' followers might have different interest tendencies as shown in Figure 1. Based on followers' interest tendencies, we could recommend suitable content to the brand for popularity enhancement. We cooperated with a fast food company from the end of 2018. We provided results of analysis to

Table 3: Spearman's ranking correlation coefficient between the proposed methods and questionnaires.

		Pointcard100 Dataset			Creditcard81 Dataset		
		Purchase	Interest	Knowledge	Purchase	Interest	Knowledge
Image	brand_img [12]	0.20	0.24	0.18	0.23	0.25	0.14
	hist-p	0.26	0.34	0.28	0.20	0.31	0.15
	hist-hi	0.26	0.38	0.33	0.23	0.32	0.17
	avg-p	0.27	0.39	0.34	0.25	0.35	0.18
Tag	hist-freq-p	0.36	**0.53**	0.49	0.31	0.40	0.24
	hist-freq-hi	0.36	**0.53**	0.49	**0.34**	0.42	**0.26**
	hist-score-p	0.33	**0.53**	**0.50**	0.31	0.40	0.22
	hist-score-hi	0.36	0.52	**0.50**	**0.34**	**0.43**	**0.26**
	avg-freq-p	0.32	0.51	0.49	0.30	0.39	0.22
	avg-score-p	0.34	0.50	0.48	0.31	0.41	0.22
Co-purchasing result obtained by point/credit card data		**0.52**	0.23	0.21	0.28	0.33	0.17

the brand including tags ranking by tag score, similar and dissimilar brands to it. And the fast food company changed their posts based on our recommendations. We recorded the number of favorites of every post from that fast food company's official account. The average number of favorites is 477 from 2015 to 2018, and increased to 3,788 in 2019. The number of favorites suddenly increased after the consulting with us proves that our proposed methods could be a novel marketing tool for popularity boosting.

7 CONCLUSIONS

In this paper, we proposed a new scale for measuring the similarity between brands using image features and tag features of social media posts made by their followers. Then, we evaluated our results by comparing with the results of the co-purchasing history of real-world customers and the results from the questionnaires. We found that our proposed methods have moderate correlations with the customers' co-purchasing histories. In addition, we found that results based on our proposed methods show stronger correlation with users' interest tendencies obtained from questionnaires than results obtained from the customers' co-purchasing histories. We believe that this multimedia-based follower analysis in social networks can become a novel marketing tool.

ACKNOWLEDGMENTS

REFERENCES

[1] Rajaraman Anand and DU Jeffrey. 2012. Mining of massive datasets.
[2] Oren Barkan and Noam Koenigstein. 2016. Item2vec: neural item embedding for collaborative filtering. In *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing*. 1–6.
[3] Tammo HA Bijmolt, Michel Wedel, Rik GM Pieters, and Wayne S DeSarbo. 1998. Judgments of brand similarity. *International Journal of Research in Marketing* 15, 3 (1998), 249–268.
[4] Kofi Boakye, Sachin Farfade, Hamid Izadinia, Yannis Kalantidis, and Pierre Garrigues. 2017. Tag prediction at Flickr: A view from the darkroom. In *Proceedings of the on Thematic Workshops of ACM Multimedia*. 376–384.
[5] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5 (2017), 135–146.
[6] John S Breese, David Heckerman, and Carl Kadie. 1998. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann Publishers Inc., 43–52.
[7] Aron Culotta and Jennifer Cutler. 2016. Mining brand perceptions from twitter social networks. *Marketing science* 35, 3 (2016), 343–362.
[8] Lisette De Vries, Sonja Gensler, and Peter SH Leeflang. 2012. Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing. *Journal of interactive marketing* 26, 2 (2012), 83–91.
[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 248–255.
[10] Ben Eisner, Tim Rocktäschel, Isabelle Augenstein, Matko Bosnjak, and Sebastian Riedel. 2016. emoji2vec: Learning Emoji Representations from their Description. In *Proceedings of the International Workshop on Natural Language Processing for Social Media*. 48–54.
[11] Reto Felix, Philipp A Rauschnabel, and Chris Hinsch. 2017. Elements of strategic social media marketing: A holistic framework. *Journal of Business Research* 70 (2017), 118–126.
[12] Francesco Gelli, Tiberio Uricchio, Xiangnan He, Alberto Del Bimbo, and Tat-Seng Chua. 2018. Beyond the Product: Discovering Image Posts for Brands in Social Media.. In *Proceedings of the ACM International Conference on Multimedia*. 465–473.
[13] Khim-Yong Goh, Cheng-Suang Heng, and Zhijie Lin. 2013. Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. *Information Systems Research* 24, 1 (2013), 88–107.
[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 770–778.
[15] Wei-Lin Hsiao and Kristen Grauman. 2017. Learning the Latent "Look": Unsupervised Discovery of a Style-Coherent Embedding from Fashion Images. In *Proceedings of the IEEE International Conference on Computer Vision*. 4213–4222.
[16] Shaowei Liu, Peng Cui, Wenwu Zhu, Shiqiang Yang, and Qi Tian. 2014. Social embedding image distance learning. In *Proceedings of the ACM International Conference on Multimedia*. 617–626.
[17] Masoud Mazloom, Robert Rietveld, Stevan Rudinac, Marcel Worring, and Willemijn Van Dolen. 2016. Multimodal popularity prediction of brand-related social media posts. In *Proceedings of the ACM International Conference on Multimedia*. 197–201.
[18] Michael J Pazzani and Daniel Billsus. 2007. Content-based recommendation systems. In *The adaptive web*. Springer, 325–341.
[19] Gerard Salton and Michael J McGill. 1986. Introduction to modern information retrieval. (1986).
[20] Cristina Segalin, Fabio Celli, Luca Polonio, Michal Kosinski, David Stillwell, Nicu Sebe, Marco Cristani, and Bruno Lepri. 2017. What your Facebook profile picture reveals about your personality. In *Proceedings of the ACM International Conference on Multimedia*. 460–468.
[21] Jizhe Wang, Pipei Huang, Huan Zhao, Zhibo Zhang, Binqiang Zhao, and Dik Lun Lee. 2018. Billion-scale commodity embedding for e-commerce recommendation in alibaba. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 839–848.
[22] Xin Xin, Xiangnan He, Yongfeng Zhang, Yongdong Zhang, and Joemon Jose. 2019. Relational Collaborative Filtering: Modeling Multiple Item Relations for Recommendation. In *Proceedings of the International ACM SIGIR Conference on Research & Development in Information Retrieval*.
[23] Hoseong Yang and Sungzoon Cho. 2015. Understanding Brands with Visualization and Keywords from eWOM using Distributed Representation. (2015).